AMENDMENTS TO THE SPECIFICATION

Please amend the paragraphs and Tables: [006], [0011], [0014], [0018], [0019], [0022], [0023], [0027], [0028], [0029], [0030], [0031], [0032], [0035], [0050], [0055], Table 1, Table 2, Table 3, Table 4, [0067], [0073], [0084], [00114], [00115], [00116], [00117], [00121], [00122], [00187], [00190], [00191], and [00192], as follows:

[006] In one embodiment, a composition is provided that comprises a protein in crystalline form wherein the protein has 65%, ~~70, 80, 90,~~ 70%, 80%, 90%, 95% or greater identity with residues ~~51-778~~ 39-766 of SEQ. ID No. 1.

[0011] In one embodiment, a method is provided for forming crystals of a protein comprising: forming a crystallization volume comprising: a protein that has at least 65%, ~~70, 80, 90,~~ 70%, 80%, 90%, 95% identity with residues ~~51-778~~ 39-766 of SEQ. ID No. 1 in a concentration between 1 mg/ml and 50 mg/ml; 5-50% w/v of precipitant wherein the precipitant comprises one or more members of the group consisting of PEG MME having a molecular weight range between 300-10000, and PEG having a molecular weight range between 100-10000; optionally 0.05 to 0.8M additives wherein the additives comprises sarcosine or 0.5 to 25% additives wherein the additives comprises xylitrol; and wherein the crystallization volume has a pH between pH 5 and pH 9; and storing the crystallization volume under conditions suitable for crystal formation. The method optionally further comprises using 0.05-0.2M buffers selected from the group consisting of Tris-HCl, bicine and combinations thereof. The method also optionally further includes performing the crystallization at a temperature between 1°C - 25°C.

[0014] In one embodiment, machine readable data storage medium is provided having data storage material encoded with machine readable data, the machine readable data comprising: structure coordinates that have a root mean square deviation of alpha-carbon atoms of less than 3Å when superimposed on alpha-carbon atoms positions of corresponding atomic coordinates of Figure 3, the root mean square deviation being calculated based only on those

alpha-carbon atoms of amino acid residues in the structure coordinates that are also present in residues ~~51-778~~ 39-766 of SEQ. ID No. 1.

[0018] In one embodiment, a method is provided for displaying a three dimensional representation of a structure of a protein comprising: taking machine readable data comprising structure coordinates that have a root mean square deviation of alpha-carbon atoms of less than 3 Å when superimposed on alpha-carbon atom positions of corresponding atomic coordinates of Figure 3, the root mean square deviation being calculated based only on those alpha-carbon atoms of amino acid residues in the structure coordinates that are also present in residues shown in Tables 1, 2, 3 and/or 4 or residues ~~51-778~~ 39-766 of SEQ. ID No. 1; computing a three dimensional representation of a structure based on the structure coordinates; and displaying the three dimensional representation.

[0019] In another embodiment, a method is provided for displaying a three dimensional representation of a structure of a protein comprising: displaying a computer model for a protein binding pocket, at least a portion of the computer model having a surface contour that has a root mean square deviation of less than 3Å when superimposed on a surface contour defined by atomic coordinates of Figure 3, the root mean square deviation being calculated based only on alpha-carbon atoms in the structure coordinates of Figure 3 that are present in residues shown in Tables 1, 2, 3 and/or 4 or residues ~~51-778~~ 39-766 of SEQ. ID No. 1.

[0022] In one embodiment, a computational method is provided comprising: taking machine readable data comprising structure coordinates that have a root mean square deviation of alpha-carbon atoms of less than 3Å when superimposed on alpha-carbon atom positions of corresponding atomic coordinates of Figure 3, the root mean square deviation being calculated based only on those alpha-carbon atoms of amino acid residues in the structure coordinates that are also present in residues shown in Tables 1, 2, 3 and/or 4 or residues ~~51-778~~ 39-766 of SEQ. ID No. 1; computing phases based on the structural coordinates; computing an electron density map based on the computed phases; and determining a three-dimensional crystal structure based on the computed electron density map.

[0023] In another embodiment, a computational method is provided comprising: taking an X-ray diffraction pattern of a crystal of the target protein; and computing a three-dimensional electron density map from the X-ray diffraction pattern by molecular replacement, wherein structure coordinates used as a molecular replacement model comprise structure coordinates that have a root mean square deviation of alpha-carbon atoms of less than 3 Å when superimposed on alpha-carbon atom positions of corresponding atomic coordinates of Figure 3, the root mean square deviation being calculated based only on those alpha-carbon atoms of amino acid residues in the structure coordinates that are also present in residues shown in Tables 1, 2, 3 and/or 4 or residues ~~51-778~~ 39-766 of SEQ. ID No. 1. This method may optionally further comprise determining a three-dimensional crystal structure based upon the computed three-dimensional electron density map.

[0027] In one embodiment, a method is provided for evaluating a potential of an entity to associate with a protein comprising: creating a computer model of a protein structure using structure coordinates that comprise structure coordinates that have a root mean square deviation of alpha-carbon atoms of less than 3 Å when superimposed on alpha-carbon atom positions of corresponding atomic coordinates of Figure 3, the root mean square deviation being calculated based only on those alpha-carbon atoms of amino acid residues in the structure coordinates that are also present in residues shown in Tables 1, 2, 3 and/or 4 or residues ~~51-778~~ 39-766 of SEQ. ID No. 1; performing a fitting operation between the entity and the computer model; and analyzing results of the fitting operation to quantify an association between the entity and the model.

[0028] In another embodiment, a method is provided for evaluating a potential of an entity to associate with a protein comprising: computing a computer model for a protein binding pocket, at least a portion of the computer model having a surface contour that has a root mean square deviation of less than 3 Å when superimposed on a surface contour defined by atomic coordinates of Figure 3, the root mean square deviation being calculated based only on alpha-carbon atoms in the structure coordinates that are present in residues shown in Tables 1, 2, 3

and/or 4 or residues ~~51-778~~ 39-766 of SEQ. ID No. 1; evaluating a potential of an entity to associate with the surface contour by performing a fitting operation between the entity and the surface contour; and analyzing results of the fitting operation to quantify an association between the entity and the computer model.

[0029] In another embodiment, a method is provided for identifying entities that can associate with a protein comprising: generating a three-dimensional structure of a protein using structure coordinates that comprise structure coordinates that have a root mean square deviation of alpha-carbon atoms of less than 3 Å when superimposed on alpha-carbon atom positions of corresponding atomic coordinates of Figure 3, the root mean square deviation being calculated based only on those alpha-carbon atoms of amino acid residues in the structure coordinates that are also present in residues shown in Tables 1, 2, 3 and/or 4 or residues ~~51-778~~ 39-766 of SEQ. ID No. 1; employing the three-dimensional structure to design or select an entity that can associate with the protein; and contacting the entity with a protein having at least 65% identity with residues ~~51-778~~ 39-766 of SEQ. ID No. 1.

[0030] In another embodiment, a method is provided for identifying entities that can associate with a protein comprising: computing a computer model for a protein binding pocket, at least a portion of the computer model having a surface contour that has a root mean square deviation of less than 3Å when superimposed on a surface contour defined by atomic coordinates of Figure 3, the root mean square deviation being calculated based only on alpha-carbon atoms in the structure coordinates that are present in residues shown in Tables 1, 2, 3 and/or 4 or residues ~~51-778~~ 39-766 of SEQ. ID No. 1; employing the computer model to design or select an entity that can associate with the protein; and contacting the entity with a protein having at least 65%, 70, 80, 90, 95% identity with residues ~~51-778~~ 39-766 of SEQ. ID No. 1.

[0031] In another embodiment, a method is provided for evaluating the ability of an entity to associate with a protein, the method comprising: constructing a computer model defined by structure coordinates that comprise structure coordinates that have a root mean square

deviation of alpha-carbon atoms of less than 3Å when superimposed on alpha-carbon atom positions of corresponding atomic coordinates of Figure 3, the root mean square deviation being calculated based only on those alpha-carbon atoms of amino acid residues in the structure coordinates that are also present in residues shown in Tables 1, 2, 3 and/or 4 or residues ~~51-778~~ 39-766 of SEQ. ID No. 1; selecting an entity to be evaluated by a method selected from the group consisting of (i) assembling molecular fragments into the entity, (ii) selecting an entity from a small molecule database, (iii) *de novo* ligand design of the entity, and (iv) modifying a known ligand for DPPIV, or a portion thereof; performing a fitting program operation between computer models of the entity to be evaluated and the binding pocket in order to provide an energy-minimized configuration of the entity in the binding pocket; and evaluating the results of the fitting operation to quantify the association between the entity and the binding pocket model in order to evaluate the ability of the entity to associate with the binding pocket.

[0032] In another embodiment, a method for evaluating the ability of an entity to associate with a protein, the method comprising: computing a computer model for a protein binding pocket, at least a portion of the computer model having a surface contour that has a root mean square deviation of less than 3Å when superimposed on a surface contour defined by atomic coordinates of Figure 3, the root mean square deviation being calculated based only on alpha-carbon atoms in the structure coordinates that are present in residues shown in Tables 1, 2, 3 and/or 4 or residues ~~51-778~~ 39-766 of SEQ. ID No. 1; selecting an entity to be evaluated by a method selected from the group consisting of (i) assembling molecular fragments into the entity, (ii) selecting an entity from a small molecule database, (iii) *de novo* ligand design of the entity, and (iv) modifying a known ligand for an DPPIV, or a portion thereof; performing a fitting program operation between computer models of the entity to be evaluated and the binding pocket in order to provide an energy-minimized configuration of the entity in the binding pocket; and evaluating the results of the fitting operation to quantify the association between the entity and the binding pocket model in order to evaluate the ability of the entity to associate with the said binding pocket.

[0035] In another embodiment, a method is provided for identifying an entity that associates with a protein comprising: taking structure coordinates from diffraction data obtained from a crystal of a protein that has at least 65%, 70%, 80%, 90%, 95% or more identity with the residues ~~51-778~~ 39-766 of SEQ. ID No. 1; and performing rational drug design using a three dimensional structure that is based on the obtained structure coordinates. The protein crystals may optionally have a crystal lattice having unit cell dimensions, +/- 5%, of a=121.53Å b=124.11Å and c=144.42Å, $\alpha=\gamma=90^{\circ}$, $\beta=114.6^{\circ}$. The method may optionally further comprise selecting one or more entities based on the rational drug design and contacting the selected entities with the protein. The method may also optionally further comprise measuring an activity of the protein when contacted with the one or more entities. The method also may optionally further comprise comparing activity of the protein in a presence of and in the absence of the one or more entities; and selecting entities where activity of the protein changes depending whether a particular entity is present. The method also may optionally further comprise contacting cells expressing the protein with the one or more entities and detecting a change in a phenotype of the cells when a particular entity is present.

[0050] In another embodiment, DPPIV comprises residues ~~51-778~~ 39-766 of SEQ. ID No. 1 which comprises the active site domain of wild-type DPPIV that is represented in the set of structural coordinates shown in Figure 3.

[0055] One or more of the sets of amino acids set forth in the tables is preferably conserved in a variant of DPPIV. Hence, DPPIV may optionally comprise a sequence that has at least 65% identity, preferably at least 70%, 80%, 90%, 95% or higher identity with any one of the above sequences (e.g., all of SEQ. ID No. 1 or residues ~~51-778~~ 39-766 of SEQ. ID No. 1) where at least the residues shown in Tables 1, 2, 3 and/or 4 are conserved with the exception of 0, 1, 2, 3, or 4 residues. It should be recognized that one might optionally vary some of the binding site residues in order to determine the effect such changes have on structure or activity.

Table 1: Amino Acids encompassed by a 4-Angstrom radius around the DPPIV active site.

ARG [[137]] 125	TYR [[559]] 547	TYR [[678]] 666
GLU [[217]] 205	SER [[642]] 630	ASN [[722]] 710
GLU [[218]] 206	TYR [[643]] 631	HIS [[752]] 740
SER [[221]] 209	VAL [[668]] 656	ASP [[720]] 708
PHE [[369]] 357	TYR [[674]] 662	

Table 2: Amino Acids encompassed by a 7-Angstrom radius around the DPPIV active site.

ARG [[137]] 125	TYR [[559]] 547	TRP [[671]] 659
HIS [[138]] 126	GLY [[561]] 549	TYR [[674]] 662
TRP [[213]] 201	PRO [[562]] 550	ASP [[675]] 663
GLU [[216]] 204	TYR [[597]] 585	TYR [[678]] 666
GLU [[217]] 205	TRP [[641]] 629	THR [[679]] 667
GLU [[218]] 206	SER [[642]] 630	ARG [[681]] 669
VAL [[219]] 207	TYR [[643]] 631	TYR [[682]] 670
PHE [[220]] 208	GLY [[644]] 632	ASN [[722]] 710
SER [[221]] 209	TYR [[646]] 634	VAL [[723]] 711
ARG [[368]] 356	ALA [[666]] 654	HIS [[752]] 740
PHE [[369]] 357	PRO [[667]] 655	ASP [[720]] 708
ARG [[370]] 358	VAL [[668]] 656	

Table 3: Amino Acids encompassed by a 10-Angstrom radius around the DPPIV active site.

ARG [[137]] 125	ILE [[417]] 405	SER [[669]] 657
HIS [[138]] 126	VAL [[558]] 546	ARG [[670]] 658
TRP [[213]] 201	TYR [[559]] 547	TRP [[671]] 659
VAL [[214]] 202	ALA [[560]] 548	TYR [[673]] 661
TYR [[215]] 203	GLY [[561]] 549	TYR [[674]] 662
GLU [[216]] 204	PRO [[562]] 550	ASP [[675]] 663
GLU [[217]] 205	CYS [[563]] 551	SER [[676]] 664
GLU [[218]] 206	SER [[564]] 552	VAL [[677]] 665
VAL [[219]] 207	TYR [[597]] 585	TYR [[678]] 666
PHE [[220]] 208	MET [[603]] 591	THR [[679]] 667
SER [[221]] 209	LEU [[610]] 598	GLU [[680]] 668
ALA [[222]] 210	GLU [[614]] 602	ARG [[681]] 669
TYR [[268]] 256	GLY [[640]] 628	TYR [[682]] 670
CYS [[313]] 301	TRP [[641]] 629	MET [[683]] 671
GLN [[332]] 320	SER [[642]] 630	HIS [[716]] 704
TRP [[365]] 353	TYR [[643]] 631	ASP [[720]] 708
VAL [[366]] 354	GLY [[644]] 632	ASP [[721]] 709
GLY [[367]] 355	GLY [[645]] 633	ASN [[722]] 710
ARG [[368]] 356	TYR [[646]] 634	VAL [[723]] 711
PHE [[369]] 357	VAL [[647]] 635	HIS [[724]] 712
ARG [[370]] 358	VAL [[665]] 653	GLN [[727]] 715
PRO [[371]] 359	ALA [[666]] 654	HIS [[752]] 740
SER [[372]] 360	PRO [[667]] 655	GLY [[753]] 741
GLU [[373]] 361	VAL [[668]] 656	

Table 4: Amino Acids encompassed by a 5-Angstrom radius around the AB and CD dimerization interfaces.

Chain A	Chain B	Chain C	Chain D
SER A [[251]] 239	PRO B [[246]] 234	PRO C [[246]] 234	LEU D [[247]] 235
TYR A [[253]] 241	ILE B [[248]] 236	LEU C [[247]] 235	ILE D [[248]] 236
SER A [[254]] 242	GLU B [[249]] 237	ILE C [[248]] 236	GLU D [[249]] 237
ASP A [[255]] 243	TYR B [[250]] 238	GLU C [[249]] 237	TYR D [[250]] 238
GLU A [[256]] 244	SER B [[251]] 239	TYR C [[250]] 238	SER D [[251]] 239
LEU A [[258]] 246	TYR B [[253]] 241	TYR C [[253]] 241	SER D [[254]] 242
GLN A [[259]] 247	SER B [[254]] 242	SER C [[254]] 242	ASP D [[255]] 243
TYR A [[260]] 248	ASP B [[255]] 243	ASP C [[255]] 243	GLU D [[256]] 244
PRO A [[261]] 249	GLU B [[256]] 244	GLU C [[256]] 244	SER D [[257]] 245
LYS A [[262]] 250	SER B [[257]] 245	SER C [[257]] 245	LEU D [[258]] 246
THR A [[263]] 251	TYR B [[260]] 248	LEU C [[258]] 246	GLN D [[259]] 247
ARG A [[265]] 253	THR B [[263]] 251	GLN C [[259]] 247	TYR D [[260]] 248
TYR A [[268]] 256	ARG B [[265]] 253	TYR C [[260]] 248	PRO D [[261]] 249
LYS A [[270]] 258	GLN B [[726]] 714	PRO C [[261]] 249	THR D [[263]] 251
ALA A [[271]] 259	ALA B [[729]] 717	THR C [[263]] 251	ARG D [[265]] 253
SER A [[732]] 720	GLN B [[730]] 718	TYR C [[268]] 256	LYS D [[270]] 258
LYS A [[733]] 721	LYS B [[733]] 721	PRO C [[269]] 257	ALA D [[271]] 259
LEU A [[735]] 723	LEU B [[735]] 723	LYS C [[270]] 258	ALA D [[273]] 261
VAL A [[736]] 724	VAL B [[736]] 724	GLU C [[672]] 660	TYR D [[673]] 661
ASP A [[737]] 725	ASP B [[737]] 725	THR C [[699]] 687	MET D [[701]] 689
GLY A [[739]] 727	GLY B [[739]] 727	LEU C [[714]] 702	HIS D [[716]] 704
VAL A [[740]] 728	VAL B [[740]] 728	PHE C [[725]] 713	GLN D [[726]] 714
PHE A [[742]] 730	ASP B [[741]] 729	GLN C [[726]] 714	SER D [[728]] 716
GLN A [[743]] 731	PHE B [[742]] 730	SER C [[728]] 716	GLN D [[730]] 718
MET A [[745]] 733	GLN B [[743]] 731	GLN C [[730]] 718	SER D [[732]] 720
TRP A [[746]] 734	ALA B [[744]] 732	LEU C [[735]] 723	ASP D [[737]] 725
TYR A [[747]] 735		VAL C [[736]] 724	VAL D [[740]] 728
THR A [[748]] 736		ASP C [[737]] 725	ASP D [[741]] 729
ASP A [[749]] 737		GLY C [[739]] 727	PHE D [[742]] 730
		VAL C [[740]] 728	GLN D [[743]] 731
		ASP C [[741]] 729	ALA D [[744]] 732
		PHE C [[742]] 730	MET D [[745]] 733
		GLN C [[743]] 731	TRP D [[746]] 734
		ALA C [[744]] 732	TYR D [[747]] 735
		MET C [[745]] 733	THR D [[748]] 736
		TRP C [[746]] 734	
		TYR C [[747]] 735	

[0067] In one variation, the variant and/or fragment of wild-type DPPIV is functional in the sense that the resulting protein is capable of associating with at least one same chemical entity that is also capable of selectively associating with a protein comprising the wild-type DPPIV (e.g., residues ~~51-778~~ 39-766 of SEQ. ID No. 1) since this common associative ability evidences that at least a portion of the native structure has been conserved. That chemical entity may optionally be glucagon-like peptide 1 (GLP-1), glucagon-like peptide 2 (GLP-2), glucose-dependent, insulinotropic polypeptide (GIP), growth hormone releasing factor, SDF-1α, β-Casomorphin, TNF-α, Peptide YY or Substance P.

[0073] The gene encoding DPPIV can be isolated from RNA, cDNA or cDNA libraries. In this case, the portion of the gene encoding amino acid residues ~~51-778~~ 39-766 (SEQ. ID No. 1), corresponding to the catalytic domain of human DPPIV, was isolated and is shown as SEQ. ID No. 2.

[0084] It should be understood that forming crystals comprising DPPIV and crystals comprising DPPIV according to the invention are not intended to be limited to the wild type, full length DPPIV shown in SEQ. ID No. 1 and to fragments comprising residues ~~51-778~~ 39-766 of SEQ. ID No. 1. Rather, it should be recognized that the invention may be extended to various other fragments and variants of wild-type DPPIV as described above.

[00114] Chain A includes amino acid residues ~~52 to 778~~ 40-766 and four amino acid residues have covalently linked sugar molecules (Figure 3). Chain B includes amino acid residues ~~47 to 778~~ 39-766 and also includes 4 histidine residues of the N-terminal polyhistidine tag (residues 35-38). Five amino acid residues of chain B have covalently linked sugar molecules (Figure 3). Chain C includes amino acid residues ~~52 to 778~~ 40-766 and five of the amino acid residues are covalently linked to sugar molecules. Chain D includes amino acid residues ~~51 to 778~~ 39-766 with five sugar-linked amino acid residues. In addition, chains C and D have no density for amino acid residues ~~151, 152 and 153~~ 139, 140, and 141 and hence coordinates for these residues are not included in Figure 3. Similarly, coordinates for amino acid residues ~~97, 98 and 99~~ 85, 86, and 87 of chain D are not included

in Figure 3. The coordinate set additionally includes 928 solvent molecules modeled as water.

[00115] Figure 4A illustrates a ribbon diagram overview of the structure of DPPIV, highlighting secondary structural elements of the protein. DPPIV is a cylindrical shaped molecule with an approximate height of 70 Å and a diameter of 60 Å (Fig. 4A). The catalytic triad of DPPIV (~~Ser642~~ Ser 630, ~~Asp720~~ Asp 708 and ~~His752~~ His 740) is illustrated in the center of Figure 4A by a "ball and stick" representation. This triad of amino acids is located in the peptidase domain or catalytic domain of DPPIV. The catalytic domain is covalently linked to the β-propeller domain (Fig. 4A).

[00116] The catalytic domain of DPPIV includes residues ~~1-67~~ 39-55 and ~~511-778~~ 499-766. Since, the structure of the present invention does not contain the first 46 residues (Chain B of Figure 3) it is presumed that the N-terminal residues of the catalytic domain adopt a random structure with a short double turn α-helix formed by residues ~~56 to 63~~ 44 to 51. The catalytic domain of DPPIV adopts a characteristic α/β hydrolase fold. The core of this domain contains an 8-stranded β-sheet with all strands being parallel except one (Fig. 4A). The β-sheet is significantly twisted and is flanked by three α-helices on one side and five α-helices on the other. The topology of the β-strands is 1, 2, -1x, 2x and (1x) (J. S. Richardson: The anatomy and taxonomy of protein structure; (1981) *Adv. Protein Chem.* **269**, 15076-15084.).

[00117] Figure 4B illustrates the remaining residues ~~68-510~~ 56-498 that form the non-catalytic domain of DPPIV. This domain is also known as β-propeller domain (Figure 4B). The β-propeller domain is a 7-fold repeat of four-stranded antiparallel β-sheets (Figure 4B). The sheets are twisted and arranged around a central tunnel as seen in case of Prolyl Oligopeptidase. Further, the β-sheets pack face-to-face and are stabilized predominantly by hydrophobic interactions. The β-propeller is linked to the catalytic domain by two polypeptide chains, one involving the N-terminal residues and the other consisting of the C-terminal residues ~~511-520~~ 499-508 which also form an α-helix.

[00121] Figure 5 illustrates the inhibitor-binding site of DPPIV based on the determined crystal structure (coordinates shown in Figure 3). The active site containing the catalytic triad (~~Ser642~~ Ser 630, ~~Asp 720~~ Asp 708 and ~~His 752~~ His 740), is located in a large cavity (Figure 5) at the interface of the catalytic and the β-propeller domains. ~~Ser 642~~ Ser 630 is located on a sharp turn that connects an α-helix to a β-strand. The positioning of this active site Serine residue is referred to as a nucleophile elbow and is characteristic of an α/β type hydrolase (D. J. Ollis et al., The α/β hydrolase fold; (1992) *Protein Eng.* **5**, 197-211). In DPPIV, the active-site serine is surrounded by hydrophobic residues, which include the large aromatic residues ~~Trp 641~~ Trp 629 and ~~Tyr 643~~ Tyr 631. The hydroxyl group of the active site serine is exposed and involved in hydrogen bonding with the imidazole group of the active site ~~His 752~~ His 740 (OH------NH distance 2.7 Å). ~~His 752~~ His 740 is located on the middle of a loop that connects a β-strand to an α-helix. The other nitrogen atom of the imidazole ring of ~~His 752~~ His 740 forms a hydrogen bond with the side chain oxygen of the third active site residue (~~Asp 720~~ Asp 708) of the catalytic triad. ~~Asp 720~~ Asp 708 is also located on a loop connecting a β-strand and an α-helix. The second oxygen atom of the side chain carboxylate of ~~Asp720~~ Asp 708 forms two hydrogen bonded interactions with the main-chain amide of residues (~~Asn722~~ Asn 710 and ~~Val723~~ Val 711). The hydrogen bonding interactions of the catalytic triad is similar to those observed for prolyl oligopeptidase.

[00122] Based on sequence alignments and structural comparisons with prolyl oligopeptidase, the residues that form the DPPIV active site pocket can be predicted with a high degree of probability. The binding pocket appears to be formed by a pocket of hydrophobic residues (~~Phe 369~~ Phe 357, ~~Tyr 643~~ Tyr 631, ~~Tyr 674~~ Tyr 662, ~~Tyr 678~~ Tyr 666, ~~Tyr 559~~ Tyr 547 and ~~Val 723~~ Val 711). In addition to the catalytic triad a large number of polar residues are also present in this hydrophobic environment (~~Arg 137~~ Arg 125, ~~Glu 217~~ Glu 205, ~~Glu 218~~ Glu 206 and ~~Asp 675~~ Asp 663).

[00187] The portion of the gene encoding residues ~~51-778~~ 39-766 (from SEQ. ID No. 1), which corresponds to the extracellular portion of human DPPIV, was isolated by PCR from spleen cDNA and cloned into the *Bam*H I and *Hind* III sites of a modified pFastBacHTb vector. This vector encodes a baculovirus glycoprotein gp67 signal peptide sequence followed by a 6x-histidine tag sequence followed by the DPPIV sequence. Expression in this vector allowed for the production of secreted recombinant DPPIV with part of a gp67 signal sequence and a 6x-histidine tag, the sequence of which is shown in Figure 1 (part of a gp67 signal sequence and 6x-histidine tag sequence underlined) (SEQ. ID No. 3).

[00190] In a typical batch prep, 5 L of cell culture supernatant was concentrated to 0.1 L on a 10 kDa NMWCO Omega Ultrasette (Pall Life Sciences) using a Masterflex L/S pump fitted with PharMed #15 tubing at a cross flow of approximately 1 L/minute and an inlet feed pressure of 1.5 to 2.0 bar, generating an initial permeate flow of up to 70 ml/minute. The retentate was diluted two to three fold by adding 25 mM Tris/HCl pH 7.9, 0.4 M NaCl and reconcentrated to 0.1 L. This process was repeated at least twice, after which the concentrate was ~~quantitavely~~ quantitatively removed from the system, centrifuged when necessary (15 minutes at 4000 rpm in an Allegra (Beckman) centrifuge) and added to approximately 8 ml of a preconditioned 50 % slurry of Probond (Invitrogen) divided over three or four 50ml conical tubes. The tubes were rotated for at least 1 hour, after which the resin was washed with 10 resin volumes of 50 mM Potassium Phosphate pH 7.9, 0.4 M NaCl, 0.25 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) ~~TCEP~~. The resin is poured into 1 cm ID glass columns (Omnifit) and washed with 50 column volumes of 50 mM Potassium Phosphate pH 7.9, 0.4 M NaCl, 20 mM imidazole, 0.25 mM TCEP. After a wash with 5 column volumes of 50 mM Tris pH 7.9, 0.4 M NaCl, 0.25 mM TCEP, the product is eluted with 4 column volumes of 50 mM Tris pH 7.9, 0.4 M NaCl, 200 mM imidazole, 0.25 mM TCEP.

[00191] It is noted that the polyhistidine tags may optionally be removed; however in this instance, the polyhistidine tag was left as a fusion. It is also noted that for the purification of non-secreted proteins, leupeptin is added to all the buffers used during the immobilized

metal affinity purification (IMAC) ~~IMAC~~ process at 1 mg/L and that for simplicity reasons the same is sometimes done when purifying DPPIV.

[00192] After concentrating to 7.5 mg/ml or higher by centrifugal ultrafiltration (10 kDa NMWCO, VivaScience), DPPIV was purified over a BioSep Sec S3000 column (200mm x 21.2 mm, Phenomenex) at 8 ml/minute to remove oligomeric forms. The column was set up in a Summit HPLC system (Dionex) managed by Chromeleon software (Dionex) and equilibrated with 25 mM Tris pH 7.6, 150 to 250mM NaCl (optionally with 0.25 mM TCEP and 1 mM EDTA). In cases when the size exclusion step was omitted, centrifugal ultrafiltration (10 kDa NMWCO) was used for the buffer exchange to the required formulation buffer. The process was carried out at 2–10°C and DPPIV was stored at the same temperature. For long-term storage, it was kept at –80°C. The purity of DPPIV was estimated by SDS-PAGE and IEF to be at least 95%. Glycosylation was confirmed by a molecular mass shift, determined by SDS-PAGE, following ~~Endo-F1 enzyme~~ treatment with endo-beta-N acetyl glucosaminidases F (Endo-F1 enzyme), and by carbohydrate analysis.